A genetic variant protective for COVID-19 is inherited from Neanderthals

Authors

Hugo Zeberg[1,2*] and Svante Pääbo[1,3*]

Affiliations

[1] Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.

[2] Department of Neuroscience, Karolinska Institutet, SE-17177 Stockholm, Sweden.

[3] Okinawa Institute of Science and Technology, Onna-son, Okinawa 904-0495, Japan.

***Corresponding authors:** hugo.zeberg@ki.se, paabo@eva.mpg.de

Abstract

It was recently shown that the major genetic risk factor associated with becoming severely ill with COVID-19 when infected by SARS-CoV-2 is inherited from Neandertals. With increasing sample size, additional genetic associations with severe COVID-19 are now being discovered. The latest study of the COVID-19 Host Genetics Initiative finds five novel genome-wide associations with COVID-19 requiring hospitalization. Here we show that a haplotype at one of these genomic regions is inherited from Neandertals. It encodes proteins that activate enzymes that are important during infections with RNA viruses. In contrast to the previously described Neandertal risk haplotype, this Neandertal haplotype is protective against severe COVID-19 but is of more moderate effect than the risk haplotype.

Main text

Neandertals and Denisovans are archaic hominin groups that became extinct about 40,000 years ago (Higham *et al*. 2014). They have a biological impact on human physiology today through genetic contributions to modern human populations that occurred during the last tens of thousands of years of their existence (*e.g*., Simonti *et al*. 2016, Dannemann and Kelso 2017). Several of these contributions affect genes involved in the immune system (*e.g*., Laurent *et al*. 2011, Quach *et al*. 2016). In particular, variants at several loci containing genes involved in innate immunity come from Neandertals and Denisovans, for example toll-like receptor gene variants which decrease the susceptibility to *Helicobacter pylori* infections and increase the risk for allergies (Dannemann *et al*. 2016).

Recently, it was shown that a region on chromosome 3 harbors the major genetic risk locus for becoming critically ill upon infection with the novel coronavirus SARS-CoV-2 (Ellinghaus *et al.* 2020) and that the haplotype that confers the risk was contributed by Neandertals to modern humans (Zeberg and Pääbo 2020). This haplotype reaches carrier frequencies of up to ~65% in South Asia whereas it is almost absent in East Asia and occurs at ~16% in Europe. Thus, although this haplotype is detrimental for its carriers during the current pandemic, it has likely been beneficial in earlier times in South Asia, perhaps by conferring protection against other pathogens.

The COVID-19 Host Genetics Initiative (HGI) is continuously adding patients in an effort to identify additional genetic factors affecting the risk of becoming infected with SARS-CoV-2 and developing COVID-19, the disease caused by the virus. In the most recent COVID-19 HGI data release, that includes 6,492 hospitalized COVID-19 patients and 1,012,809 controls, five new loci with genome-wide significant effects were identified in addition to the risk locus on chromosome 3 (https://www.covid19hg.org/). Two of these are located on chromosome 6, and one each on chromosomes 12, 19 and 21 (Fig. 1). To test if any of these loci carry variants derived from Neandertals we investigated if any of the single nucleotide polymorphisms (SNPs) with genome-wide significance in these regions carried alleles that matched three high-quality Neandertals genomes, while being absent in the genomes of 108 African Yoruba individuals. No SNPs fulfilling these criteria were found in four of the novel loci. In contrast, the locus on chromosome 12 harbors 18 such SNPs.

We next scrutinized 35 SNPs with genome-wide significance which have been scored in the archaic genomes at the chromosome 12 locus (Table S1). A 60,000 to 80,000-year-old Neandertal from Chayrskaya Cave in southern Siberia (Mafessoni *et al*. 2020) carries the protective alleles at 34 of these SNPs in a homozygous form, while an approximately 50,000-year-old Neandertal from Vindija in Croatia

(Prüfer *et al*. 2017) carries the protective alleles in a homozygous form in 33 cases. An approximately 120,000-year-old Neandertal from Denisova Cave in southern Siberia (Prüfer *et al*. 2014) carries protective alleles in homozygous form at 30 of the SNPs and in a heterozygous form at three of the SNPs. In contrast, only two of the protective alleles match the Denisovan genome (Meyer *et al*. 2012). Thus, for almost all SNPs with genome-wide significance on chromosome 12, the Neandertals carry the allele that decrease the risk of requiring hospitalization upon SARS-CoV-2 infection. For the variant with the largest association (p=5.7e-9, rs4766664) the reduction in risk is ~14% (under the rare disease assumption, OR = 0.86, 95% confidence interval (CI): 0.82-0.91).

By identifying sets of Neandertals alleles that co-segregate in humans living today, we find a Neandertal-like segment of ~90 kilobases (kb) (chr12:113,345,168-113,435,449; *hg19*) which coincide with genomic region harboring SNPs associated with severe COVID-19 (Fig 2.; Methods). Genomic segments with similarity to Neandertal genomes may either derive from common ancestors of the two groups living about half a million years ago or be contributed by Neandertals to modern humans when the two groups met less than 100,000 years ago (Sankararaman *et al*. 2012). To distinguish these two alternatives, we use the recombination rate of the genomic region (0.80cM/Mb) (Kong *et al*. 2002), a generation time of 29 years (Langergraber *et al*. 2012), a split between Neandertals and modern humans of 550,000 years (Prüfer *et al*. 2014), interbreeding between the two groups ~50,000 years ago, and a published equation (Huerta-Sánchez *et al*. 2014) to test if a segment of this length may have survived since the common ancestor of the groups without being broken down by recombination that affect chromosomes in each generation. Under these assumptions, we exclude that it derives from the common ancestor (p = 1.1e-10) and conclude that this region entered the human gene pool from Neandertals. In agreement with this, a previous study (Mendez *et al*., 2013) has described gene flow from Neandertals in this region. We find that the index allele, *i.e*. the Neandertal allele with the greatest association with severe COVID-19 (rs4766664), occurs in all populations in Eurasia and Americas at frequencies that often reach and exceed 30% (Fig. 3).

The Neandertal haplotype protective against severe COVID-19 is located in a region in chromosome 12 that contain the three genes *OAS1, OAS2* and *OAS3*, all of which encode oligoadenylate synthetases. These enzymes are induced by interferons and activated by double-stranded RNA. They produce short-chain polyadenylates, which in turn activate ribonuclease L, which degrades intracellular RNA, and activate other antiviral mechanisms in infected cells (reviewed by Choi *et al.* 2015).

To investigate which of these genes might be involved in the protection against severe COVID-19, we plot the genomic location of the *OAS* genes below the p-values for the association and the introgressed

Neandertal haplotype (Fig. 2). The association (p<1e-6) overlaps all three *OAS* genes. The SNPs with the most significant associations (p<5e-8) are situated in the 3'-end of *OAS1* and in *OAS3*, and thus these two genes are most likely to be involved in the protective effect(s) of the Neandertal haplotype.

There are several SNPs on the Neandertal haplotype which stand out as potentially functionally important in this region. The SNP which has attracted most attention (rs10774671) affects a splice acceptor site (Sams *et al*. 2016). The most common allele at this SNP is derived and alters splicing of the *OAS1* transcript that produces an ancestral protein isoform (p46) to several alternative transcripts (Li *et al*. 2017). The ancestral isoform, preserved in Neanderthals, has been shown to have a higher enzymatic activity (Bonnevie-Nielsen *et al*. 2005). Outside Africa the ancestral allele is only present in individuals with the Neandertal haplotype whereas in Africa it exists presumably as a genetic variant inherited from the common ancestors of modern humans and Neandertals that was lost in the modern human population that left Africa (Sams *et al*. 2016). We note that a previous study (Carey *et al*. 2019) has identified recurrent *OAS1* loss-of-function mutations in primates, suggesting a cost of OAS1 enzymatic activity. In addition to the splice acceptor site, the 90kb Neandertal haplotype contains a missense variant (rs2660) in one *OAS1* transcript, a missense variant (rs1859330) and two synonymous variants (rs1859329, rs2285932) in *OAS3*, and a missense variant in *OAS2* (rs1293767).

The haplotype on chromosome 12 has previously been studied with respect to its effects in other viral infections. Notably, the Neandertal missense variant in *OAS1* (or variants in linkage disequilibrium with this variant) has been shown to be associated with moderate to strong protection against SARS-CoV (rs2660, OR = 0.42, 95% CI: 0.20-0.89) (He *et al*. 2006), although this study was limited in number of SARS-CoV cases and controls. The Neandertal-like splice acceptor variant has similarly been associated with protection against West Nile Virus (rs10774671, OR = 0.63, 95% CI: 0.5-0.83, Kim *et al*. 2009) and the Neandertal-like haplotypes have been associated with increased resistance to hepatitis C infections (El Awady *et al*., 2011). Furthermore, the Neandertal versions of the *OAS* genes have been shown to be expressed differently in response to different viral infections both in terms of expression levels and splice forms (Sams *et al*., 2016). Thus, it is entirely possible that the *OAS* locus has been subject to distinct selective pressures by different pathogens during the hundreds of thousands of years that Neandertals and the ancestors of modern humans lived in western Eurasia and Subsaharan Africa, respectively.

After the groups came into contact and fused ~50,000 years ago, they became exposed to novel pathogens. Some of the variants contributed by Neandertals rose to high frequencies either locally, as is the case in South Asia for the Neandertal haplotype on chromosome 3 (Zeberg and Pääbo 2020), or almost everywhere in Eurasia, as is the case for the Neandertal haplotype on chromosome 12 (Fig. 3).

When almost the entire world population is now exposed to the novel SARS-CoV-2, the former haplotypes turn out to be detrimental while the latter haplotype turns out to be beneficial. Unfortunately, the protection against severe disease conferred by the Neandertal *OAS* variant is substantially smaller than the increased risk conferred by the chromosome 3 factor.

Methods

The summary statistics used here relies on the round 4 (alpha) release of the meta-analysis carried out by the COVID-19 Host Genetics Initiative (https://covid19hg.org/results). Linkage disequilibrium was calculated using LDlink 4.1 and alleles were compared to the archaic genomes using tabix (HTSlib 1.10). The Neandertal haplotype covering the region was investigated using sites where the Neandertal allele is not present in 108 Yoruba individuals (1000G Genomes Project), and the distances between the first two SNPs which cause the LD to fall below 0.75 in Europeans was taken as the length of the Neandertal haplotype. The probability of observing a haplotype of a certain length or longer due to incomplete lineage sorting was calculated as described (Huerta-Sánchez *et al.* 2014). The inferred ancestral states at variable positions among present-day humans were taken from Ensembl. Maps displaying allele frequencies of different populations were made using Mathematica 11.0 (Wolfram Research, Inc., Champaign, IL) and OpenStreetMap data.

Acknowledgments. We are indebted to the COVID-19 Host Genetics Initiative (HGI) for making the GWAS data available and to the Max Planck Society and the NOMIS Foundation for funding.

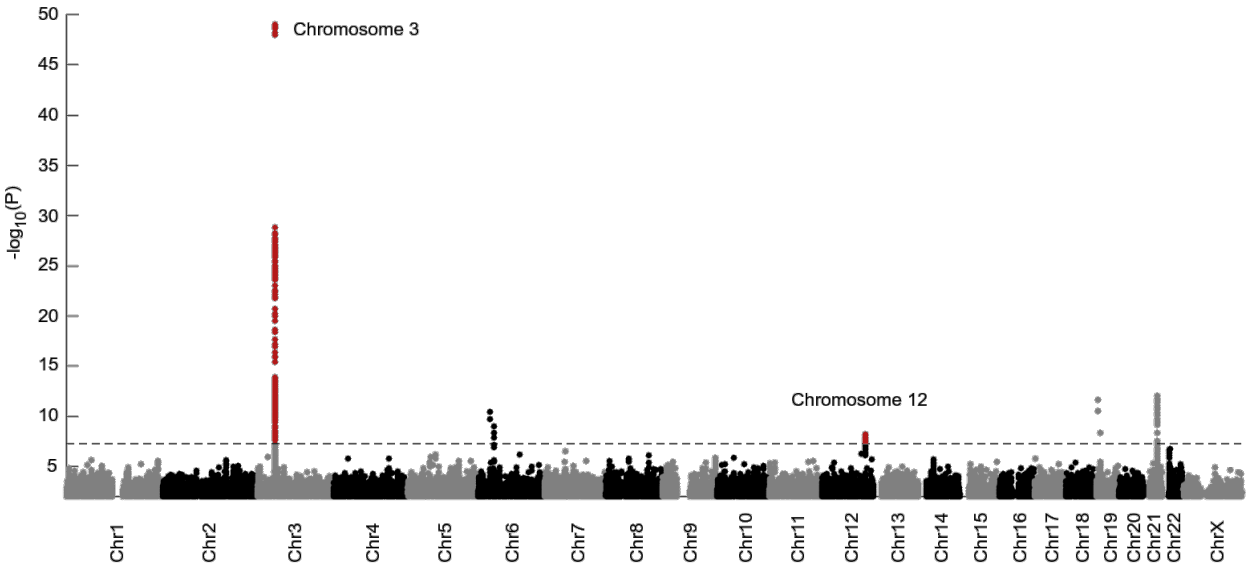


Figure 1. Genetic variants associated with severe COVID-19. Out of six genomic regions with genome-wide significance (dashed line, p = 5e-8), two carry alleles that match Neandertal genomes (red). The previously identified risk factor on chromosome 3 has the strongest association with severe COVID-19 (p=9.5e-50, index variant rs35081325), whereas the protective variant on chromosome 12 is less significantly associated (p=5.7e-9, index variant rs4766664).

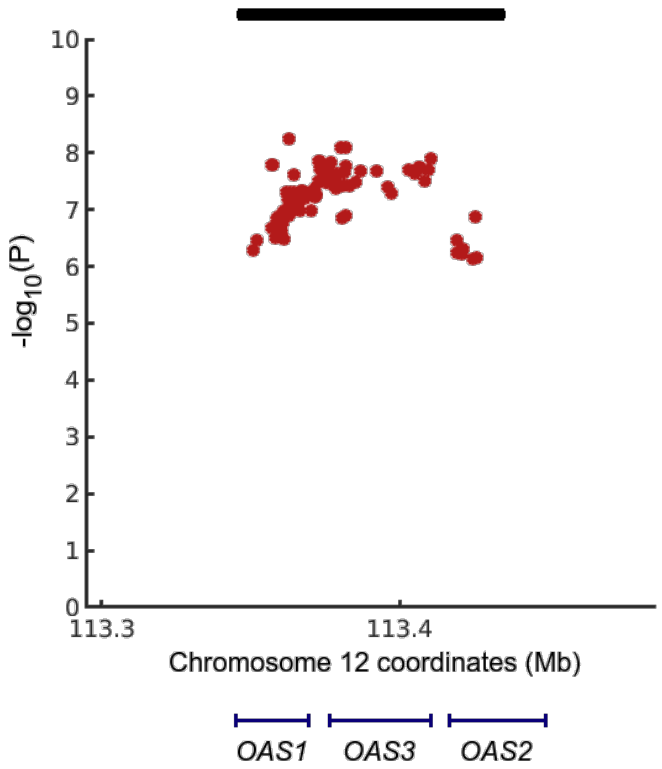


Figure 2. Genetic variants associated with COVID-19 hospitalization (p<1e-6) at the *OAS* locus. Black bar indicates a Neanderal haplotype defined by 66 diagnostic Neandertal variants in linkage disequilibrium in Europeans (r^2>0.75, Methods, Fig. S1). Blue lines indicate the three *OAS* genes, all of which are transcribed from left to right. The *X*-axis gives *hg19* coordinates.

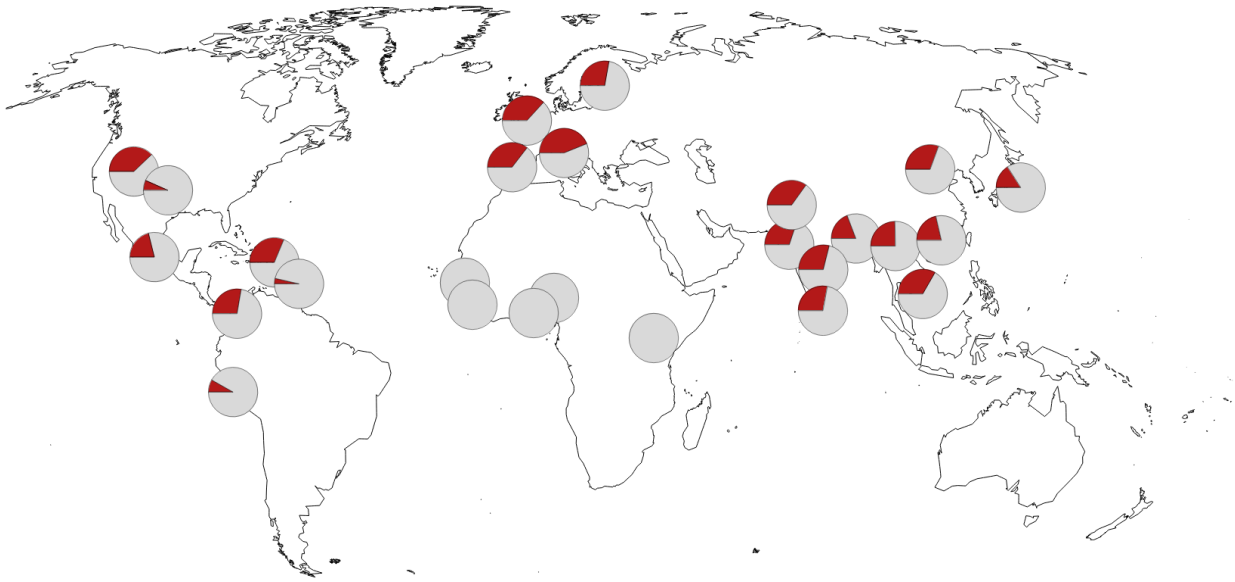


Figure 3. Geographic distribution of the allele indicative of the Neandertal haplotype protective against severe COVID-19. Pie charts indicate minor allele frequency at rs4766664. Frequency data from the 1000 Genomes Project. Map source data from OpenStreetMap.

References

1. Bonnevie-Nielsen, V. *et al*. (2005) Variation in antiviral 2-prime,5-prime-oligoadenylate synthetase (2-prime5-prime AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am. J. Hum. Genet*. doi: 10.1086/429391

2. Carey, C.M. *et al*. (2019) Recurrent loss-of-function mutations reveal costs to OAS1 antiviral activity in primates. *Cell Host Microbe*. doi: 10.1016/j.chom.2019.01.001

3. Choi, U.Y. *et al.* (2015) Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp. Mol. Med*. doi: 10.1038/emm.2014.110

4. Dannemann, M. *et al*. (2016) Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* doi: 10.1016/j.ajhg.2015.11.015

5. Dannemann, M. and Kelso, J. (2017) The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *Am. J. Hum. Genet.* doi: 10.1016/j.ajhg.2017.09.010

6. El Awady, M.K. *et al*. (2011) Single nucleotide polymorphism at exon 7 splice acceptor site of OAS1 gene determines response of hepatitis C virus patients to interferon therapy. *J. Gastroenterol. Hepatol*. doi: 10.1111/j.1440-1746.2010.06605.x

7. Ellinghaus, D. et al. (2020) Genome-wide Association Study of Severe Covid-19 with Respiratory Failure. *NEJM.* doi: 10.1056/NEJMoa2020283

8. He, J. *et al.* (2020) Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case-control study. *BMC Infectious Diseases*. doi: 10.1186/1471-2334-6-106

9. Higham, T. *et al*. (2014) The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*. doi: 10.1038/nature13621

10. Huerta-Sánchez, E., *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*. doi: 10.1038/nature13408

11. Kim, JK. *et al*. (2009) Genetic Variation in OAS1 Is a Risk Factor for Initial Infection with West Nile Virus in Man. *PLOS Pathogens*. doi: 10.1371/journal.ppat.1000321

12. Kong, A. *et al*. (2002). A high-resolution recombination map of the human genome. *Nat. Genet*. doi: 10.1038/ng917

13. Langergraber, K.E. *et al*. (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U.S.A*. doi: 10.1073/pnas.1211740109

14. Laurent, A. *et al*. (2011) The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. *Science*. doi: 10.1126/science.1209202

15. Li, H. *et al*. (2017) Identification of a Sjögren's syndrome susceptibility locus at OAS1 that influences isoform switching, protein expression, and responsiveness to type I interferons. *PLOS Genetics*. doi: 10.1371/journal.pgen.1006820

16. Mafessoni, F. *et al*. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci. U.S.A.* doi: 10.1073/pnas.2004944117

17. Meyer, M. *et al*. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science*. doi: 10.1126/science.1224344

18. Prüfer, K. *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. doi: 10.1038/nature12886

19. Prüfer, K. *et al*. (2017) A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. doi: 10.1126/science.aao1887

20. Quach, H. *et al*. (2016). Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell*. doi: 10.1016/j.cell.2016.09.02410.1016/j.cell.2016.09.024

21. Sams, A.J. *et al*. (2016) Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biol*. doi: 10.1186/s13059-016-1098-6.

22. Sankararaman, S. *et al*. (2012). The date of interbreeding between Neandertals and modern humans. *PLoS Genet*. doi: 10.1371/journal.pgen.1002947

23. Simonti, C.N. *et al*. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science*. doi: 10.1126/science.aad2149

24. Zeberg, H. and Pääbo, S. (2020) The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. doi: 10.1038/s41586-020-2818-3.

Supplementary material

Table S1. Genome-wide significant variants ($p < 5e-8$) on chromosome 12.

Figure S1. Diagnostic Neandertal variants and their pairwise linkage disequilibrium in Europeans.

Chr	Pos	rsid	P-value	LD (rs4766664)	Ref	Alt	Anc	Vindija	Altai	Chagyrskaya	Denisova
12	113362997	rs4766664	5.7e-09	1.00	T	G	T	T/T	T/T	T/T	G/G
12	113381956	rs2269899	7.9e-09	0.56	C	T	T	C/C	C/C	C/C	T/T
12	113380008	rs10735079	7.9e-09	0.75	G	A	A	G/G	G/G	G/G	A/A
12	113410316	rs10744791	1.3e-08	0.63	G	A	A	G/G	A/G	G/G	A/A
12	113372804	rs4767037	1.4e-08	0.98	A	C	A	A/A	A/A	A/A	C/C
12	113372866	rs1981557	1.4e-08	0.98	G	C	G	G/G	G/G	G/G	C/C
12	113372977	rs1981555	1.4e-08	0.98	G	A	G	.	.	.	.
12	113376913	rs7299132	1.5e-08	0.99	T	A	A	T/T	T/T	T/T	A/A
12	113357209	rs1131476	1.6e-08	0.92	G	A	G	G/G	G/G	G/G	A/A
12	113357442	rs2660	1.6e-08	0.92	G	A	G	G/G	G/G	G/G	A/A
12	113381695	rs7131998	1.8e-08	0.86	A	C	A	.	.	.	.
12	113406196	rs3937434	1.8e-08	0.63	A	G	G	.	.	.	.
12	113406460	rs2016831	1.8e-08	0.63	G	C	C	.	.	.	.
12	113374017	rs4767040	1.9e-08	0.98	G	C	G	G/G	G/G	G/G	C/C
12	113372961	rs1981556	1.9e-08	1.00	C	G	G	.	.	.	.
12	113406945	rs757405	1.9e-08	0.63	T	A	A	T/T	A/T	T/T	A/A
12	113373563	rs3759376	1.9e-08	1.00	A	G	G	A/A	A/A	A/A	G/G
12	113409413	rs4767045	1.9e-08	0.63	C	T	T	T/T	T/T	C/C	T/T
12	113402899	rs4767044	2.0e-08	0.63	C	A	A	C/C	C/C	C/C	A/A
12	113373588	rs3759375	2.0e-08	1.00	A	G	G	A/A	A/A	A/A	G/G
12	113386950	rs2285932	2.0e-08	0.63	T	C	C	T/T	T/T	T/T	C/C
12	113375036	rs7132797	2.1e-08	0.98	A	C	A	A/A	A/A	A/A	C/C
12	113392182	rs7310667	2.1e-08	0.31	A	G	A	.	.	.	.
12	113381376	rs6489882	2.2e-08	0.97	G	A	A	A/A	A/A	A/A	A/A
12	113377822	rs6489879	2.3e-08	0.99	G	A	G	G/G	G/G	G/G	A/A
12	113379123	rs7311182	2.3e-08	0.99	T	C	C	.	.	.	.
12	113405181	rs1557866	2.4e-08	0.63	A	C	C	A/A	A/A	A/A	C/C
12	113380271	rs6489880	2.4e-08	0.99	C	T	T	.	.	.	.
12	113364382	rs4767033	2.4e-08	1.00	T	A	A	T/T	T/T	T/T	A/A
12	113378081	rs4238033	2.4e-08	0.99	T	A	A	T/T	T/T	T/T	A/A
12	113375983	rs1156361	2.6e-08	1.00	T	C	C	T/T	T/T	T/T	C/C
12	113372539	rs7966314	3.1e-08	0.98	A	G	A	A/A	A/A	A/A	G/G
12	113408208	rs2010604	3.2e-08	0.52	G	C	C	G/G	G/G	G/G	G/G
12	113374748	rs10774679	3.3e-08	0.44	C	T	C	C/C	C/C	C/C	C/C
12	113385000	rs10850103	3.3e-08	0.63	A	T	T	A/A	T/A	A/A	T/T
12	113376320	rs3815178	3.3e-08	0.99	C	T	C	C/C	C/C	C/C	T/T
12	113376388	rs1859330	3.4e-08	0.53	G	A	G	G/G	G/G	G/G	A/A
12	113381217	rs6489881	3.6e-08	0.75	A	T	A	A/A	A/A	A/A	T/T
12	113380708	rs7977345	3.7e-08	0.75	A	T	T	A/A	A/A	A/A	T/T
12	113379039	rs7955267	3.7e-08	0.75	C	T	C	.	.	.	.
12	113382977	rs2384074	3.7e-08	0.66	C	T	C	.	.	.	.
12	113396010	rs10744789	4.1e-08	0.35	T	C	C	.	.	.	.
12	113371646	rs9971885	4.2e-08	1.00	A	C	C	.	.	.	.
12	113378677	rs4767041	4.3e-08	0.75	G	A	G	.	.	.	.
12	113367422	rs2384072	4.6e-08	1.00	T	C	T	T/T	T/T	T/T	C/C
12	113366899	rs7306205	4.6e-08	1.00	A	G	A	A/A	A/A	A/A	G/G
12	113367309	rs1859336	4.7e-08	1.00	C	T	C	C/C	C/C	C/C	T/T
12	113364633	rs4766667	4.8e-08	1.00	G	A	G	.	.	.	.
12	113370966	rs1859333	4.9e-08	1.00	T	C	C	T/T	T/T	T/T	C/C
12	113368079	rs6489877	4.9e-08	0.99	A	G	A	.	.	.	.
12	113363408	rs6489866	4.9e-08	1.00	A	G	A	.	.	.	.

Table S1. Genome-wide significant variants (p < 5e-8) on chromosome 12. Red marks the minor allele which in all cases shown is protective. Note that the human reference genome carries the Neandertal haplotype. *'Ref', 'Alt',* and *'Anc'* refer to the reference, alternative and ancestral allele, respectively. LD denotes linkage disequilibrium in the 1000 Genomes Project dataset (r^2).

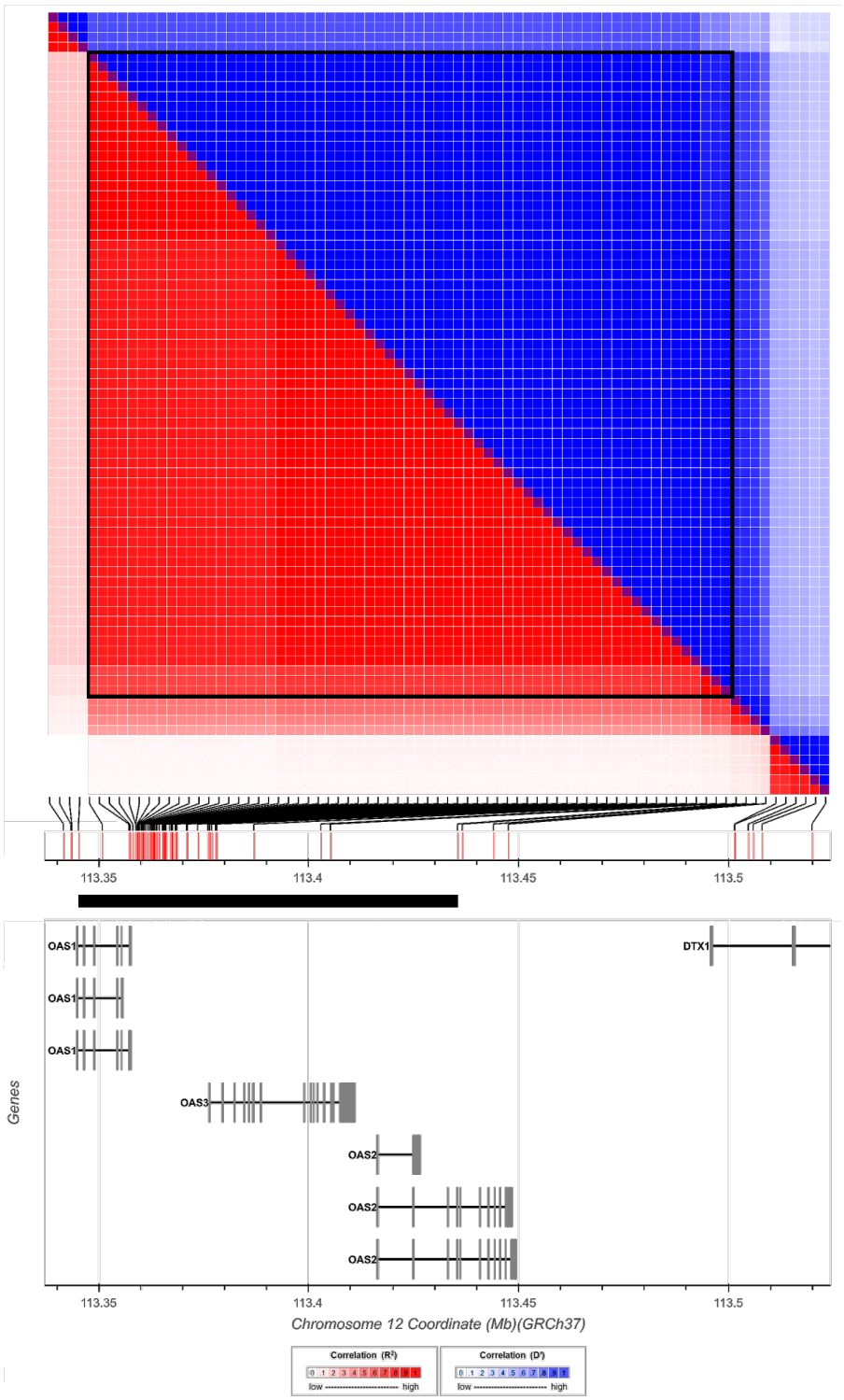
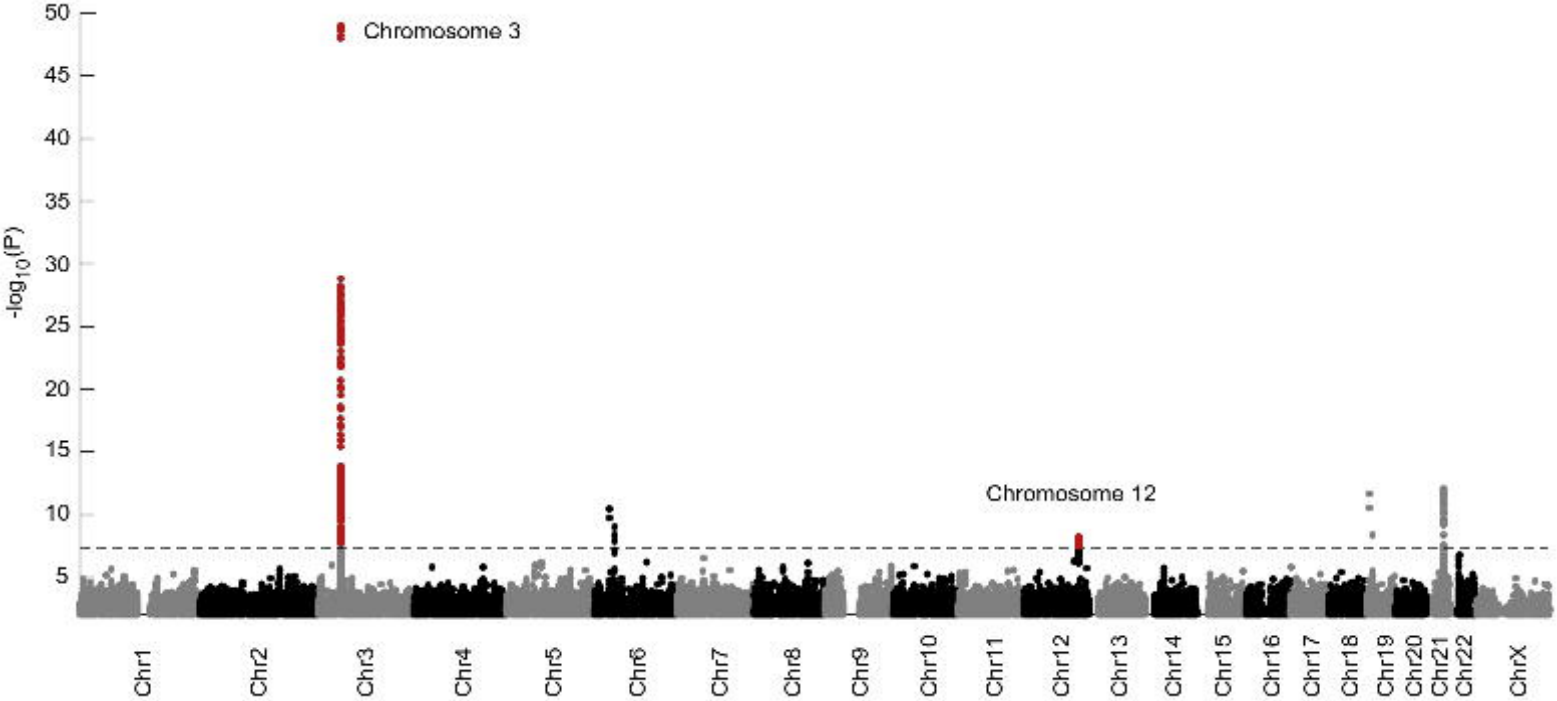
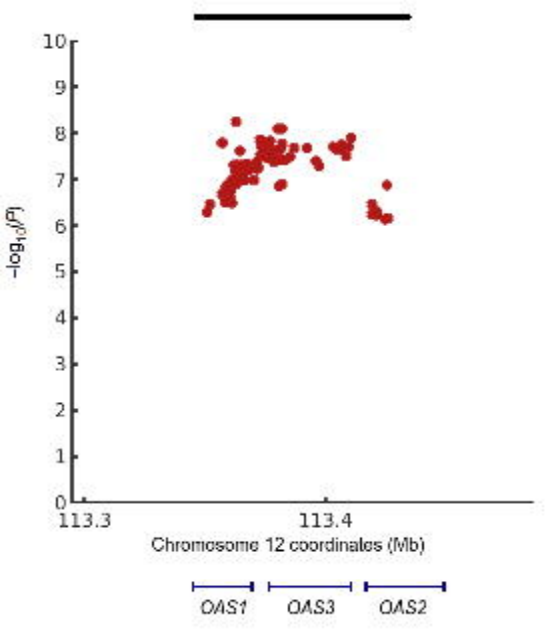
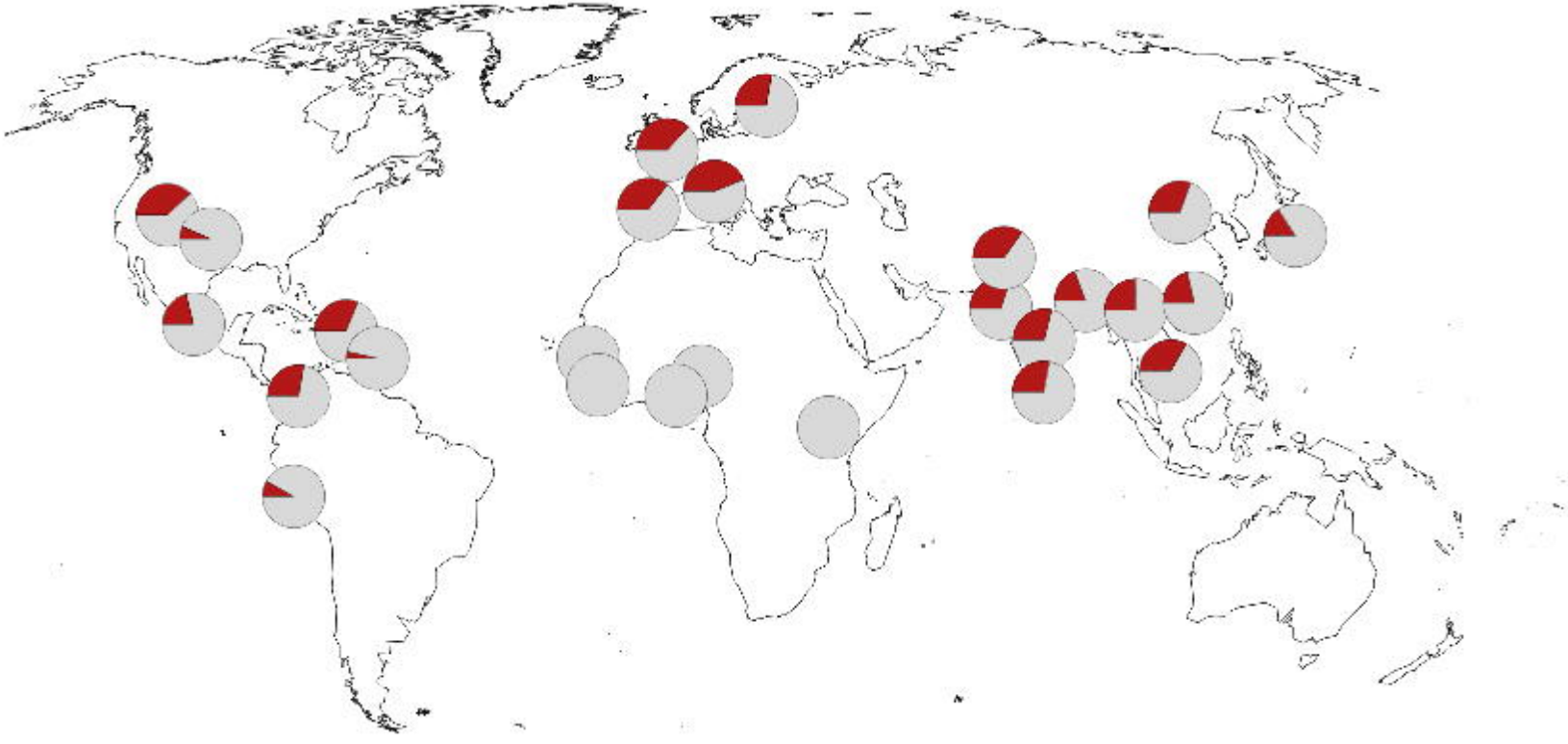


Figure S1. Diagnostic Neandertal variants and their pairwise linkage disequilibrium in Europeans. Black box indicates a set of 66 diagnostic variants in linkage disequilibrium (r^2>0.75) defining a haplotype (black bar) with coordinates chr12:113,345,168-113,435,449 (*hg19*).


Chromosome 3
Chromosome 12
-log10(P)
50
45
40
35
30
25
20
15
10
5
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX


10
9
8
7
6
5
4
3
2
1
0
$-\log_{10}(P)$
113.3
113.4
Chromosome 12 coordinates (Mb)
OAS1
OAS3
OAS2